Universal Journal of Public Health 12(1): 136-149, 2024 http://www.hrpub.org
DOI: 10.13189/ujph.2024.120115

Analysis of Schoolgirls' Mental Health – A Machine Learning Approach to Distinguish Between Academic and Abusive Stress

Deivanai Gurusamy[1], Midhunchakkaravarthy Janarthanan[1,*], Prasun Chakrabarti[1,2]

[1]Faculty of Computer Science and Multimedia, Lincoln University College, Malaysia
[2]Department of Computer Science and Engineering, Sir Padampat Singhania University, India

Received November 28, 2023; Revised January 8, 2024; Accepted February 17, 2024

Cite This Paper in the Following Citation Styles

***(a):** [1] Deivanai Gurusamy, Midhunchakkaravarthy Janarthanan, Prasun Chakrabarti , "Analysis of Schoolgirls' Mental Health – A Machine Learning Approach to Distinguish Between Academic and Abusive Stress," Universal Journal of Public Health, Vol. 12, No. 1, pp. 136 - 149, 2024. DOI: 10.13189/ujph.2024.120115.*

***(b):** Deivanai Gurusamy, Midhunchakkaravarthy Janarthanan, Prasun Chakrabarti (2024). Analysis of Schoolgirls' Mental Health – A Machine Learning Approach to Distinguish Between Academic and Abusive Stress. Universal Journal of Public Health, 12(1), 136 - 149. DOI: 10.13189/ujph.2024.120115.*

Abstract The prevalence of stress and depression is rising among students. Current research acknowledges that academic burden is a source of stress for students and evaluates physiological parameters during academic activities by deliberately introducing stressors or stress. Nonetheless, schoolgirls perceive sexual, physical, and emotional abuse as the primary source of stress nowadays. This study gathered information from fifty schoolgirls via an in-house oral questionnaire to identify those concerned about academic pressure and maltreatment. This questionnaire was designed to disclose the psychological state through oral, behavioural, and physiological responses without instigating stress unnaturally. The prepared dataset was experimented with using multinomial logistic regression and decision tree using C4.8, Naive Bayes, and k-NN for three class classifications (normal, academic, and abuse stress). The weighted average F1-score of the leading models from each classifier was 89.9%, 89.2%, 89.5%, and 88.6%, respectively. The results indicate that logistic regression performs marginally better than other methods. When separating normal and academic stress samples, the same classifier achieves an F1-score of 94.9% and 69.0%, respectively. In identifying the abuse stress classes, k-NN achieved the maximum F1-score of 88.8%. In addition, the normal, academic, and abuse stress classes achieved sensitivities of 98.4%, 65.4%, and 98.4% and specificities of 71.7%, 98.4%, and 100%, respectively. The categorisation models constructed from the research can identify schoolgirls with internalised conflict for earlier intervention.

Keywords Stress, Abuse, Wearables, In-House Questionnaire, Schoolgirls, Classification, Machine Learning

1. Introduction

Everyone encounters the challenge of stress, which results in physiological, behavioural, emotional, and structural changes [1]. There are two types of stress: acute, which lasts for a short period, and chronic, which lasts longer [2]. The stressor usually determines the severity of the stress. This stressor is a threat to a creature, and the reaction to the threat is the stress response generated by the central nervous system [3]. Stressors may be, for instance, poverty, unemployment, work pressure, family troubles, marriage problems, study load, war conditions, parental concerns, social media, and harassment [2]–[6]. When stressed, changes will occur in the individual's neurological, cardiovascular, immunological, and endocrine systems. These changes subsequently produce the stress response, which releases the hormones as energy. This energy is then delivered to the organs via increased blood pressure (BP) and heart rate [3], thus causing a rapid heart rate (HR) and perspiration in stressed people [7]. Moreover, the intense and enduring stressors may result in physical and mental health problems [3], [8].

Globally, students experience stress and, for instance, relationship troubles [2], [9], financial concerns [2], [9], [10], fear about their future and profession [9], [11], family issues [2], [10], [11], change of surroundings [12], social media [13], [14], COVID pandemic [15]–[17], academic burden [2], [9], [11], [16], [18]–[22], primarily written [23]–[26] and oral examinations [27], research defence [28], training [29], are all regarded to be the contributors for student's stress and mental health concerns. However, numerous studies [9], [11]–[13], [30], [31] cite academic burden as one of the causes of stress affecting students' mental health, and some studies [2], [12], [32]–[37] cite it as the primary cause.

Generally, researchers adopt a few techniques to obtain participant data to detect academic and other forms of stress. As machine learning (ML) techniques have gained popularity in the healthcare industry [38], data are analysed using ML to classify stressed individuals. The literature on the classification of student stress via ML raised some concerns among the authors of this study. Prior research has primarily focused on undergraduate and graduate students, with no work on the perspectives of school students' stress. Schoolgirls, in particular, incur tremendous stress, and the number of girls suffering from poor mental health is increasing [38]–[42]. Furthermore, research indicates that female students are more stressed than male students [12] [43]–[47]. Therefore, stress evaluations of schoolgirls using machine learning are necessary.

Second, studies that built an ML model to detect student stress and depression focused on academic stress. Still, TV, print, and web publications reveal many sorts of school kid maltreatment. Researchers also report that abuse and harassment are prevalent among female pupils [48]–[52]. Since sexual and other abuse is seldom reported, researchers underestimate its role in student stress and suicidal thoughts. Schoolgirls subjected to such pressure develop anxiety, which manifests physically as stress and eventually becomes internalised. This internalised personality of schoolgirls affects their academic performance, renders them unsociable with family and friends, and drives some to suicide [9], [41], [42], [48], [52]–[54]. Thus, recognising sexual and other abuse as a source of student stress is essential for early detection and treatment.

Third, the researchers used non-natural stress-inducing procedures; however, how closely these techniques match genuine stress for stress detection is unknown. Lastly, behavioural assessment has been neglected in the existing research to acquire data for stress evaluation. Personality psychology recommends observing a person's behaviour in order to determine who he or she is. No single source (wearable devices, self-report) provides a transparent view of a person's personality [55]. Therefore, the inclusion of behavioural information may improve the classifier's efficacy. These considerations prompted the authors to make a novel effort to collect behavioural data from the schoolgirls through observation and physiological data using wearable devices in real-time. At the same time, the girls orally answered a questionnaire designed to assess academic and abuse stress with the greatest sensitivity possible. The main objective of the research is to assist the schools, parents, and professionals in determining whether schoolgirls are stressed academically or due to abuse.

The paper's organisation includes the following sections: Section 2 describes the related work. Section 3 covers study methodology, and data collection, analysis. Section 4 discusses findings, while Section 5 concludes the paper.

2. Related Work

This section summarises the procedures used by researchers in identifying or classifying stressed or depressed students using ML techniques. The summary includes the tools, stimuli, wearable devices, research procedures, participants, and parameters used in the existing work, along with the results and gaps.

In [23], EEG characteristics were extracted when students saw the Chinese face system, which includes images of humans expressing different emotions. Beck Depression Inventory (BDI) Scale labels were used to label students a few days before the exam. In experiments employing the collected data, (k- Nearest Neighbor) k-NN surpassed RF (random forest), naïve Bayes (NB), support vector machine (SVM), and logistic regression (LR) with 99.1% accuracy. However, image-induced stress must be better defined. Another study [20] examined electrodermal activity (EDA) and skin temperature (ST) in undergraduates wearing the Affectiva Q sensor. Participants' phone usage, location shifts, and survey data were collected over 30 days. The participant's happiness, vitality, health, alertness, and stress were examined, and the ensemble classifier performed well with 70% accuracy. However, phone usage statistics may not be able to predict happiness in all people.

Interestingly, [27] evaluated ultra-short-term heart rate variability (HRV) from 42 university students using ECG. Data was collected during students' oral exams to measure stress. Stress and rest conditions were classified using AdaBoost (AB), SVM, NB, C4.5, and multilayer perceptron. The decision tree (DT) using C4.5 had the best area under the curve with 79% accuracy. The tiny sample size and risks of ultra-short-term HRV in real-world situations are drawbacks of this technique. Blood volume pulse data with Empatica E4 was gathered from the postgraduate students who were examined if they were stressed during defences with and without an audience [28]. The extreme learning machine (ELM) classifier classified the labelled data based on student self-report with 91.93 % accuracy. The research found that students felt agitated during defence with an audience; however, this stress is typical and does not last. Another study [18] collected data from college students during stressful tasks, including the ice cube exam, singing, and arithmetic, as well as non-stressful ones like emailing, eating, and relaxing. Their main goal was to see if wrist-worn devices outperformed chest-worn devices. RF performed well with an F1-score of 88.8% for the data. The lab study used artificial stress induction methods, which may not apply to real-world situations.

A fake snake test [56] produced stress and measured accelerometer, gyro, and angle data to diagnose internalised illnesses in youngsters. The carers completed the questionnaire to help researchers diagnose the children. Data-derived features classified children with 75% accuracy using k-NN. However, the study used many characteristics, and artificial stress inducement doubts the model's accuracy. RF and neural network methods predicted exam completion time as a stress metric from students' mouse-clicking behaviour [25]. However, other methods may have been employed besides mouse-clicking. [22] developed a self-designed biosensor to identify high-performing, low-stress kids. They collected 25 students' HR, pulse rate, activity, time, and outcomes. The k-means approach identified stress-management-ready kids. However, the self-designed sensor's reliability and accuracy needed an explanation. [24]'s data set was based entirely on students' self-reports of mood, sentiments, and past episodes, despite the authors' claim that it investigated exam dread and internet use. The stressed pupils were recognised using RF, NB, SVM, and k-NN. SVM had the highest accuracy at 85.71%; however, the research protocol needed improvement.

During their Empatica E4 training simulation, a data set was created from medical students with ST, HRV, BP, and EDA [29]. The relaxation and training simulation phases laid the groundwork for data processing. Using the collected attributes, the study used k-means to cluster students with 70% accuracy. This study cannot be used for students from various fields of study. [26] prepared two sets of exam data from two times separated by four months. HRV data from a Bio-beam device and student questionnaire data were collected. Long Short-term Memory Network (LSMN) demonstrated 83% accuracy in binary classification. However, the study is confined to HRV data. [21] examined college students' stress levels after the semester using mobile phones and Fitbit data on sleep and activity count. The ground truth for depression analysis was the BDI scale. Association rule mining extracted context-based information to identify sad students. This research uses only one modality. [19] collected undergraduate students' galvanic skin response (GSR), respiratory rate, body temperature, and pulse oximeter data using custom wearable devices. They employed physiological data to split students into stress and non-stress groups by artificially creating stress with an arithmetic problem. The experiment used SVM, k-NN, LR, and RF classifiers with different signal features. SVM was a decent classifier with 98.89% accuracy, but the study only employed one non-natural stress-generating task.

To measure 25 university students' GSR, an electronic nose [16] with gas sensors was constructed. The primary purpose was to measure students' tension during virtual exams. The GSR data classification accuracy was 100%. In [57], SVM beat decision trees, NB, and k-NN in classifying students' stress levels as normal, moderate, or severe. However, these works [16], [57] focused on young people's attitudes during COVID, which we have previously passed. The pupils' stress was assessed using a questionnaire [58], and RF performed well on the acquired data with 89.72 % accuracy. However, the study was based only on self-report. In the study [59], wearable smartwatches were used to develop a method using a discrete wavelet transform technique to achieve a balance between high and low-frequency ECG signals. While the investigation was limited to non-natural stresses, the naive bayes obtained 96.67% accuracy. There have also been some studies that used frameworks like deep learning [13] and natural language processing [60] to analyse student stress with social media as a cause.

Through a study of the relevant literature, it was determined that there are only a few works that evaluate student stress using machine learning. To the author's knowledge, recent articles have not addressed the issue since the end of the COVID-19 pandemic. In addition, there is no work dealing with the stress of female school pupils using machine learning, and most of the research has yet to focus on the actual stress experienced by students. Also, existing research has yet to utilise behavioural observation, except for [22], which analysed behaviour solely through mouse tapping patterns.

3. Materials and Methods

The entire procedure began with open-ended interviews with psychological experts. The information obtained from the expertise and experience of specialists was used to develop a questionnaire with separate sections to measure academic and abuse-related stress. Perceived stress scales [61] [62], [63] were also utilised to determine the types of questions to be included and to compute the results. Afterwards, wearable gadgets such as Fitbit Sense 2 and blood pressure monitoring devices for measuring the physiological features of the schoolgirls were selected.

Participants in the study were fifty female high school students. The selection criteria were female students in the 10th, 11th, and 12th classes aged 14 to 16 who were in good health and willing to engage in the study. They were chosen as subjects because various studies show that female students are more stressed than male students [43]–[47], notably, female school students experience abuse or harassment [48]–[52], both in and out of school. Also, there is no existing work on analysing schoolgirls' stress using machine learning. This investigation used students from diverse schools and geographic locations. Informed consent was received from the female students and their parents.

Three individuals responsible for observing behavioural changes were trained to collect the data. Instead of intentionally producing stress or detecting physiological changes during academic pursuits, the questionnaires were delivered to the individuals via a chat at their site. Physiological data such as HR and BP were recorded throughout the talk, as these parameters help more to recognise mental disorders [64], coupled with behavioural traits. The verbal responses provided by the individuals were also captured. The data-collection process attempted to follow the psychiatrist's approach by inquiring about the sentiments or ideas of individuals regarding the specific topic. Moreover, as the psychiatrist received the reaction and examined the patient's behaviour, this study monitored the patient's behaviour and recorded the response. The data collection process was carried out during three sessions with separate questionnaires: baseline, academic stress evaluation and abuse stress evaluation.

It was discovered through data investigation that physiological measurements and values derived from observation would identify the condition of stress in students. All other personal details collected were deemed to have a negligible influence on the outcome. As a result, they were all eliminated. Heart rate, systolic BP, and diastolic BP are numerical data, whereas nervous state and sweat nature are categorical data. Apart from heart rate, all other data were measured once during a session and thoroughly checked for mistakes. The mean value obtained from the heart rate data measured in each session was used to replace the missing values for each student individually.

The data analysis findings indicate that physiological readings and self-report scores correlate with Pearson's correlation coefficient value 0.61. So, the ground truth was derived from the self-report score to determine the stress state.

3.1. Feature Selection and Preparation of Datasets

Several datasets and classifiers were tested to determine which attribute best classifies occurrences into three classes. The whole dataset included mean heart rate, systolic and diastolic BP, nervous state, and perspiration state, from which further datasets were produced. For determining each modality's contribution, the datasets were manually derived: dataset 1 had only the Fitbit device's heart rate measurements; dataset 2 had only blood pressure (systolic and diastolic) readings from the blood pressure monitor; and dataset 3 had only observation-derived characteristics (nervous state and sweat nature). In addition, the Weka software's feature selection approaches were utilised. The method for selecting features includes an attribute evaluator and a search method.

The attribute evaluator evaluates each feature in the context of the output class, and the search method tries various feature combinations to find the optimal combination for accurately predicting the class. This experiment analysed evaluators' performance, such as wrapper subset evaluation (WSE), correlation-based feature selection (CFS) subset evaluation, correlation attribute evaluation, and information gain attribute evaluation, using ten-fold cross-validation on the complete data set that included samples of all three classes. Search methods such as Best First (BF) and Ranker were utilised based on the evaluator's requirements.

The WSE [65] evaluates various feature subsets selected by the BF search method based on the learning algorithm's accuracy. The CFS subset evaluation [65] evaluates the subset of attributes chosen by the BF method based on the ability of each feature to predict and the degree of redundancy among the subset's features. The correlation attribute and the information gain attribute methods evaluate each attribute by quantifying Pearson's coefficient and the information gain concerning the attribute and the class, respectively. Both utilise a ranker search method that ranks each attribute according to its evaluability [65].

Table 1 summarises the results of the feature selection methods. The WSE for classifier k-NN and Naïve Bayes resulted in the subset with the same features; the other two have distinct subsets. The best subset of heart rate, diastolic BP, and nervousness was found with the merit of 44% by the CFS subset evaluation method. However, the WSE has identified the subsets with the highest merit, with an average of 90.2% across all classifiers. Dataset 4 was derived from the features selected by the WSE for each classifier, and dataset 5 was from the CFS subset method. In correlation attribute-based evaluation methods, the features were ranked in the order of heart rate, nervous, systolic BP, diastolic BP, and sweat, based on the correlation between the feature and the class. The first two features from the order correlate with the class more than the average (close to 1), and the third, fourth and fifth features have shallow correlation values close to 0. Heart rate is highly correlated with 62%, whereas perspiration correlation is low at 15%. Hence, the first two ranked features were selected.

Table 1. Summary of feature selection techniques

Evaluator	Search Method	Search Direction	Subsets Evaluated	Features selected	The merit of the best subset found
WSE (LR)	Best First	Backward	25	Heart rate, Diastolic BP	0.905
WSE (NB)	Best First	Backward	22	Heart rate, Nervous	0.907
WSE (C4.8)	Best First	Backward	18	Heart rate, Diastolic BP Systolic BP	0.893
WSE (k-NN)	Best First	Backward	24	Heart rate, Nervous,	0.901
CFS Subset	Best First	Backward	21	Heart rate, Diastolic BP, Nervous	0.444
Correlation Attribute	Ranker	-	-	**Heart rate (0.62), Nervous (0.57),** Systolic BP (0.21), Diastolic BP (0.19), sweat (0.154)	-
Information Gain Attribute	Ranker	-	-	**Heart rate (0.402), Nervous (0.238),** Diastolic BP (0.066), sweat (0.026), Systolic BP (0.1)	-

In the information gain attribute method, the information gain of each feature concerning the class variable is low for all; however, considering the reasonable values of heart rate and nervousness were selected, the remaining attributes that have values very close to 0 were removed. Through the results, based on correlation and information gain, the same features were selected, and that formed dataset 6. The overall results from the feature selection method imply that the heart rate and nervous state are the most contributing features for predicting the stress class.

3.2. Classification Approach

In this study, classification was carried out using multinomial LR, a decision tree constructed with C4.8, naive bayes, and k-NN classifiers in Weka 3.9.5. Multinomial LR predicts the probabilities of the classes, and NB is also a probability-based method that handles all the features equally. A tree classifier creates a hierarchical model by selecting an attribute for a root node and then extending the branches of attributes from that node. Weka uses the information gain factor to select the root attribute. In this study, the C4.8 decision tree classifier, an updated version of C4.5, was used. For classification, the k-NN depends on the distance between the test and training samples. Initially, these classifiers were evaluated with Weka's default settings, and then a few parameters were altered to determine the performance variation. If the classifier's performance was inferior, the cost-sensitive classifier was also investigated to identify the best model for the dataset. The cost-sensitive classifier allocates varying costs to the training samples, reducing misclassification errors in highly skewed classes. Randomisation was performed on the dataset containing normal, AC_Stress (academic stress), and AB_Stress (abuse stress) samples before they were input into the classifiers. The classifiers were trained and evaluated using ten-fold cross-validation, in which the entire dataset is arbitrarily divided into ten segments, of which nine out of ten are training samples. One is the test sample, and the procedure is repeated ten times. The weighted average value of the F1-score, precision, recall, specificity, receiver operating characteristics (ROC), and kappa were the metrics used for evaluation. Due to the imbalanced state of the dataset, the F1-score was used to select the most effective classifier. In cases of equal F1 scores, the F1 scores for each class were utilised.

4. Results and Analysis

Weka was used to test the multinomial LR with default parameters. The results are displayed in Table 2. The original dataset with all features yielded 90.5% precision, 90.7% recall, and 89.9% F1-score. These values are marginally more significant than the datasets 1, 4, 5 and 6. Concerning ROC, dataset 1, containing only the mean HR, has achieved 88.5%, followed by the original set with a 0.4% difference. Dataset 3, containing characteristics derived from observation, did not yield desirable results. A cost-sensitive approach was employed to improve the classifier's performance on the specific dataset by increasing the penalty for misclassified samples; however, the performance did not improve. It demonstrates that perspiration and nervous characteristics alone are insufficient for distinguishing between stress states. The F1-score results indicate that the multinomial LR performs well on the contribution of all collected features, followed by dataset 5 containing HR, diastolic BP, and nervous state values selected by the CFS method. Further, the F1-score of individual classes indicates that, except for dataset 3, all others have significantly recognised the three distinct stress states. The class 'Normal' has received the most significant samples, producing excellent F1-score values. While the class 'AC_Stress' results are also satisfactory, the class 'AB_Stress' has received the fewest samples, and its results fall between 40% and 50%. Nonetheless, dataset 1 identifies the 'AB_Stress' state with an F1-score of 80%.

Table 2. Results from multinomial LR

Dataset	Precision (Weighted avg.)	Recall (Weighted avg.)	F1-Score				ROC (Weighted avg.)
			Weighted Avg.	*Normal*	*AC_Stress*	*AB_Stress*	
1 (Modality 1)	0.899	0.900	0.889	0.945	0.634	0.800	0.885
2 (Modality 2)	0.667	0.800	0.728	0.888	0.000	0.400	0.624
3 (Modality 3)	-	0.873	-	0.932	0.596	-	0.692
4 (WSE)	0.898	0.900	0.891	0.945	0.667	0.500	0.883
5 (CFS)	0.899	0.900	0.895	0.948	0.682	0.400	0.874
6 (Correlation or Information Gain)	0.894	0.900	0.893	0.948	0.667	0.500	0.876
Original set (Modality1, 2 and 3)	0.905	0.907	0.899	0.949	0.698	0.500	0.881

Table 3. Results from the decision tree using C4.8

Dataset	Precision (Weighted avg.)	Recall (Weighted avg.)	F1-Score				ROC (Weighted avg.)
			Weighted Avg.	*Normal*	*AC_Stress*	*AB_Stress*	
1 (Modality 1)	0.892	0.893	0.892	0.939	0.680	0.800	0.754
2 (Modality 2)	0.800	0.433	0.463	0.485	0.375	0.286	0.574
3 (Modality 3)	-	0.873	-	0.929	0.609	-	0.672
4 (WSE)	0.872	0.880	0.864	0.934	0.550	0.667	0.682
5 (CFS)	0.887	0.893	0.882	0.945	0.619	0.500	0.711
6 (Correlation or Information Gain)	0.880	0.887	0.873	0.938	0.585	0.667	0.716
Original set (Modality1, 2 and 3)	0.803	0.733	0.757	0.827	0.451	0.500	0.766

Table 4. Results from naïve Bayes

Dataset	Precision (Weighted avg.)	Recall (Weighted avg.)	F1-Score				ROC (Weighted avg.)
			Weighted Avg.	*Normal*	*AC_Stress*	*AB_Stress*	
1 (Modality 1)	-	0.900	-	0.949	0.651	-	0.892
2 (Modality 2)	0.656	0.780	0.713	0.876	0.000	0.000	0.557
3 (Modality 3)	0.872	0.880	0.873	0.932	0.609	0.667	0.699
4 (WSE)	0.895	0.900	0.895	0.944	0.681	0.667	0.884
5 (CFS)	-	0.900	-	0.948	0.681	-	0.881
6 (Correlation or Information Gain)	0.895	0.900	0.895	0.944	0.681	0.667	0.884
Original set (Modality1, 2 and 3)	-	0.880	-	0.935	0.640	-	0.880

Table 5. Results from k-NN

Dataset	Precision (Weighted avg.)	Recall (Weighted avg.)	F1-Score				ROC (Weighted avg.)
			Weighted Avg.	*Normal*	*AC_Stress*	*AB_Stress*	
1 (Modality 1)	0.888	0.893	0.886	0.940	0.636	0.800	0.845
2 (Modality 2)	0.696	0.700	0.886	0.816	0.196	0.000	0.531
3 (Modality 3)	-	0.873	0.698	0.929	0.609	0.000	0.699
4 (WSE)	0.912	0.913	-	0.953	0.698	0.667	0.835
5 (CFS)	0.869	0.873	0.869	0.927	0.612	0.667	0.831
6 (Correlation or Information Gain)	0.912	0.913	0.869	0.953	0.698	0.667	0.835
Original set (Modality1, 2 and 3)	-	0.847	0.905	0.914	0.582	-	0.752

Next, the DT was evaluated using C4.8 with the default parameters and by adjusting the minimal number of instances per leaf from 2 to 20 and the confidence factor from 0.25 to 1. In addition, the concept of pruning in decision trees, which optimises leaf nodes to reduce overfitting, was enabled. Even though multiple trees with varying parameters were constructed, the models developed yielded no acceptable results except the one from dataset 1. Consequently, the base classifier was developed as a cost-sensitive classifier by experimenting with various cost matrices. The optimal results were achieved with the cost matrix [0, 1, 1; 10, 0, 1; 1, 10, 0], as shown in Table 3.

Dataset 1, containing only HR values, outperforms all other datasets regarding F1-score, recall, and precision with 89.2%, 89.3%, and 89.2%, respectively. In addition, the F1-score of individual classes is significantly greater than that of other datasets. Similarly, to the multinomial LR classifier, this classifier's performance on dataset 3 is inferior. It trails the initial set by only 1.2% regarding ROC value. 4, 5, and 6 are the datasets that generated favourable findings in addition to dataset 1. These data sets contain features chosen by feature selection techniques. The finding from DT concludes that the classifier performs well with only HR values, followed by dataset 5 with HR, diastolic BP, and nervous state selected by the CFS method.

The results of the NB classifier are presented in Table 4. The classifier with default parameters performed reasonably well on dataset 3, with an F1-score of 87.3%, a recall of 88%, a precision of 87.2%, and a ROC of 69.9%. The classifier's results on dataset 2 were also considered good, but it performed poorly in classifying abuse stress classes. The classifying model on other datasets yielded poor results; therefore, the cost-sensitive naive bayes was tested to improve the performance of other datasets; however, only datasets 4 and 6 (which contain the same features) performed better with the cost matrix [0, 1, 1; 1, 0, 1; 20, 1, 0]. Even though the F1-score and ROC values of the cost-sensitive model derived from data sets 4 and 6 are more significant than the base classifier's results on dataset 3, the cost of the models is compared to select the superior model. The average cost from the cost-sensitive model is 0.1, while the average cost from the base classifier model is 0.12, with a minor difference between them. Therefore, naive bayes yields significant results for datasets 3, 4, and 6. It is shown that the NB performs better on datasets containing features derived from WSE, correlation, and information gain methods and on datasets containing behavioural information.

The k-NN classifier was evaluated with various k values and default parameter values; Table 5 displays the results. The k value of 1 produced the best results across all datasets. Regarding the highest possible F1-score value, the original set has attained 90.5%. However, the model needed to classify 'AB_Stress' classes effectively. The models on datasets 1 and 2 have the second-highest F1 score. However, the model on dataset 1 outperforms the model on dataset 2, with significant differences in the precision, recall, ROC, and F1-Score of individual classes. Consequently, the optimal k-NN classifier model is derived from dataset 1, with only HR contributing to such results.

4.1. Comparative Analysis of Various Classifiers

After defining the best-performing model for each classifier, the best-performing classifier needed to be determined. A comparison was performed to determine the best classifier and contributing dataset using the weighted average of evaluation metric values; the resulting graph is depicted in Fig. 1. Multinomial LR has the highest F1-score value at 89.9%, followed by NB, DT, and k-NN by 0.4%, 0.7%, and 1.3%, respectively. The dataset containing all the features (mean heart rate, diastolic blood pressure, perspiration, and nervousness) yielded the highest F1-score value. These characteristics were derived from all three modalities, demonstrating the contribution of each modality to high performance. In addition, the datasets that contribute to class prediction must be reliable. In this study, the term "reliable" refers to the relevant data that the dataset contains to predict various stress classes and the extent to which the classifier makes use of such data to classify correctly.

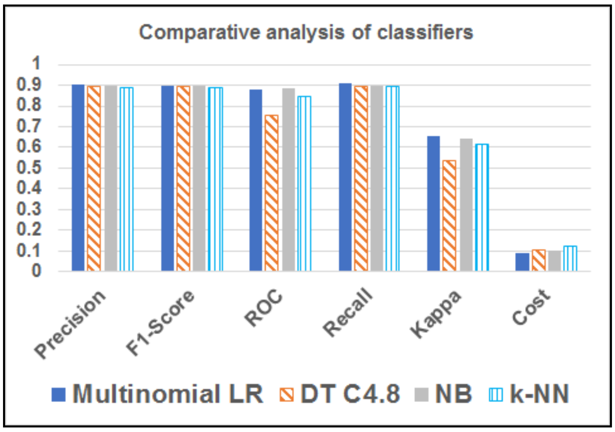

Figure 1. Comparative analysis of classifiers

The purpose of using kappa statistics and the range of permissible values is context-dependent. According to [66], it identifies the correlation between dependent and independent variables. Some research [19] uses the kappa value and the range defined by Cohen to estimate the classifier's effectiveness and the data's contribution to the best classification, even though Cohen's range is used to evaluate the reliability of data collected by multiple data collectors. In this investigation, kappa statistic is used to determine the classification accuracy of the stress classifier. According to Fig. 1, multinomial LR obtains the highest kappa statistic with a value of 0.652, and the classifier produced these results with the fewest samples. The larger the examples, the greater the kappa statistic value [66], so the classifier was retested with the dataset using the synthetic minority oversampling technique (SMOTE) that increases the number of samples in a balanced manner, yielding a kappa statistic of 0.78. According to the interpretation of Cohen's kappa value, a value between 0.61 and 0.80 indicates data reliability between 64 and 81% [19]. Therefore, the derived kappa statistic indicates that the classifier can use the dataset to predict stress levels.

The result of the NB is derived from a cost-sensitive model; consequently, the cost of constructing models with each classifier is also compared, and the multinomial LR has the lowest cost. Regarding precision, F1-score, ROC, and sensitivity, there is only a slight variation between models, indicating that all classifiers perform well on the given data set despite LR's slight lead.

To further comprehend the performance of the various classifiers in predicting the classes, the sensitivity and specificity of each class were computed, and the results are shown in Table 6. The analysis revealed that multinomial LR, DT, and DT obtained the highest sensitivity of 98.4% for the 'normal' class, 65.4% for the 'AC_Stress' class, and 100% for the 'AB_Stress' class, respectively. DT, LR and both NB and k-NN obtain the highest specificity of 71.7% for the 'normal' class, 98.4% for the 'AC_Stress' class, and 100% for the 'AB_Stress' class, respectively. Therefore, DT outperforms others in sensitivity; implementing DT as cost-sensitive could be a possible explanation. Regarding specificity, each classifier contributes to the highest value of each class.

4.2. Comparative Analysis of Various Modalities

Fig. 2 compares the weighted average F1-score of the developed models of the various classifiers with the features from the various modalities. Modality 1 (Fitbit wristwatch) contributes significantly more to the models developed by LR, DT C4.8, and k-NN than to the model developed by NB. The contribution of Modality 2 (blood pressure monitor) is present in all classifiers; however, the F1-score values vary from classifier to classifier. DT C4.8's value is relatively low, above average in LR and NB, and high in k-NN. Modality 3 (behavioural observation) did not contribute to the LR and DT scores (C4.8), but the F1-score in NB and k-NN is relatively high. Nonetheless, the best multinomial LR and NB classifying models use all three modalities, whereas DT uses only modality one, and k-NN uses modalities 1 and 3. Modality 1 contributes the most to developing the best models for all classifiers, followed by modality 3, which contributes to developing the best models for two classifiers, and modality 2, which contributes alongside other modalities to develop the best model for one classifier.

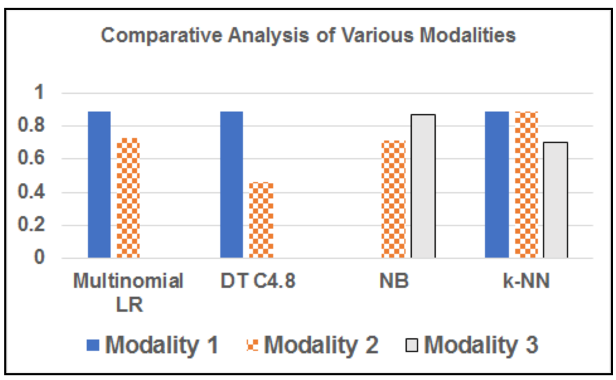

Figure 2. Comparative analysis of modalities

Table 6. Analysis of sensitivity and specificity

Classifier	Sensitivity			Specificity		
	Normal	**AC_Stress**	**AB_Stress**	**Normal**	**AC_Stress**	**AB_Stress**
Multinomial LR	0.984	0.577	0.500	0.607	0.984	0.993
DT (C4.8)	0.943	0.654	1	0.714	0.944	0.993
NB	0.967	0.615	0.500	0.643	0.960	1
k-NN	0.959	0.615	0.500	0.607	0.935	1

4.3. Analysis of Academic and Abuse Stress Results

Fig. 3 compares classifiers and outcomes regarding individual F1-score values for each class. Because the "normal" class has the most samples, it was classified with high F1-score values and an average of 94% of all classifiers. k-NN recognises abuse stress samples with a high F1-score of 88.7%, followed by C4.8. All of the classifiers earned an average F1-score of 71%. In identifying academic stress samples, LR had the greatest F1-score, followed by C4.8, NB, and k-NN, with an average value of 67%. The results reveal that all classifiers could differentiate between academic and abuse stress samples from the given dataset despite the abuse stress samples being rather small.

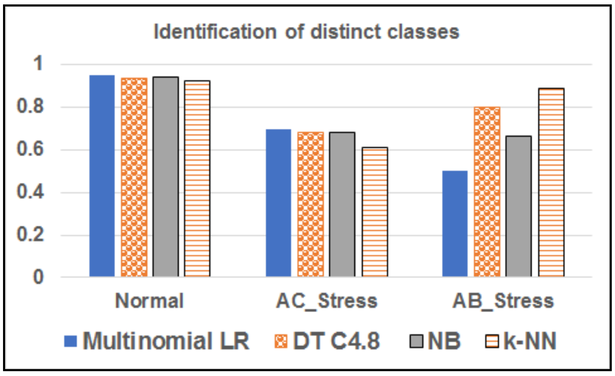

Figure 3. The effect of classifiers on class identification

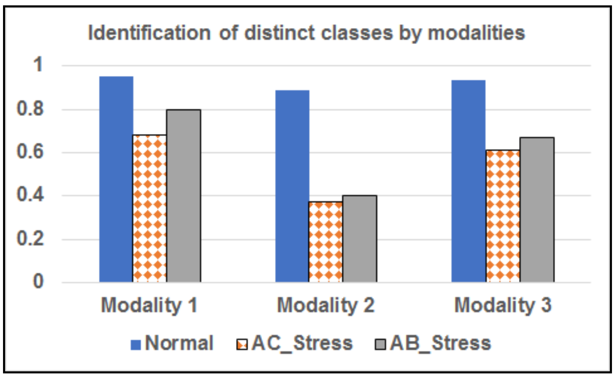

Figure 4. The Effect of modalities on class identification

Fig. 4 depicts the role of modalities in distinguishing academic and abuse stress samples. According to the data, modality 1 distinguished the classes with the highest F1-score values, followed by modality three and modality 2. According to the results, all modalities could distinguish between distinct classes.

4.4. Analysis of the Behavioural Observation Data's Results

Behavioural observation data was deemed essential to this study, so perspiration type and nervousness were included in the data collection. The feature selection techniques employed in the study selected the nervous state as one of the most crucial classification characteristics. The perspiration state was not selected and may not distinguish between the classes. During data collection, those who reported feeling tensed did not perspire but appeared nervous. Nevertheless, Fig. 4 demonstrates that behavioural observation was a second factor in identifying the stress classes 'Normal' (0.932), 'AC_Stress' (0.609), and 'AB_stress' (0.667) F1-score values. In addition, the study's most influential models incorporate behavioural data. Consequently, the results indicate that observing behaviour and adding observed data to electronically collected data will enhance the classifier's performance.

4.5. Discussion

Today's schoolchildren face numerous obstacles, which result in mental health issues. It is widely known that schoolgirls commit suicide, primarily through television, newspapers, and word of mouth. Initially, the public cited academic pressure and other factors as the cause of schoolgirls' suicide, but the media is now emphasising the role of harassment and sexual abuse. The media depicts the victims' parents stating that their daughters were normal before their suicides and friends stating that the victims were not as social as they once were. This issue requires everyone's attention. Stress awareness should be taught to schoolgirls, and institutions or parents should provide periodic counselling. Despite the schoolgirls' reluctance to express their emotional state to others, they must be able to communicate and comprehend their mental health. As a modest contribution to the schoolgirl community, this study aimed to identify stressed schoolgirls using machine learning models. To accomplish the task, an in-house questionnaire was developed to assess academic and abuse-related stress based on the knowledge gained from psychological experts and the literature concerning perceived stress scales. Instead of obtaining a written self-report, this study collected the participants' responses to the queries through a conversation.

Following the recommendations of the psychological specialists, the research aimed to use simple wearables to measure their heart rate and blood pressure. Some students may not convey their emotions through self-reporting, so the authors decided to measure physiological changes during a conversation using wearable devices. However, school students were reluctant to wear chest-worn devices or devices with threatening appearances, so simple and familiar devices were used.

The traditional method of monitoring the participants' behaviour during their conversation about the specific context was also employed to gather additional information. The study protocol was developed, and data from fifty schoolgirls was collected. After data collection, participant feedback was obtained to determine if the participants were comfortable wearing the devices and were not constrained while answering the questions. Unsurprisingly, the feedback revealed that every student felt comfortable and that the procedures were satisfactory. Therefore, the acceptance rate for the data collection method and apparatus is one hundred per cent.

However, the collected data could not express the specified range of heart rate and blood pressure values to identify stressed pupils. This prompted the authors to consider the self-report as a ground truth for analysing the stress classes, which led to the creation of a dataset containing independent variables and the dependent variable. The number of datasets derived from the complete dataset depends on the data acquisition modality and feature selection methods. According to the utilised feature selection methods, the heart rate and nervous state characteristics gleaned from behavioural observation were the most important features for classification. The classifiers LR and decision tree using C4.8, NB, and k-NN were tested on the prepared datasets. Although evaluation metrics such as precision, recall, and ROC were assessed, the F1-score was deemed the best indicator of a model and the modality that contributed the most to distinguishing academic and abuse stress.

The result of the experiment signifies that the best models were created using multinomial logistic regression, decision trees with C4.8, naive bayes, and k-NN, with respective weighted average F1-scores of 89.9%, 89.2%, 89.5%, and 88.9%. Due to its 89.9% F1-score, 89.9% precision, 90% recall, 88.5% ROC, and 65% kappa statistic derived from all the features, multinomial LR has designated the best classifier among these models. It demonstrates that all the features have contributed to this significant outcome. The sensitivity for the 'Normal' class was determined to be 98.4%, 65.4% for the 'AC_Stress' class, and 100% for the 'AB_Stress' class. The 'Normal' class had the highest specificity at 71.7%, followed by the 'AC_Stress' class at 98.4% and the 'AB_Stress' class at 100%. All classifiers recognised the 'Normal' class samples with an average F1-score of 94%. LR identified academic stress samples with the highest F1-score of 69.8%, whereas k-NN identified abuse stress samples with the highest F1-score of 88.7%.

Among the three modalities used, the Fitbit heart rate value contributes the most to classification with a high F1-score value. Behavioural information contributes the second most to the classification after wristwatch data, followed by blood pressure monitor data. It demonstrates that even though medical-grade devices and sensors aid in healthcare regarding stress analysis, a psychological analysis, the traditional method of observing behaviour, becomes essential.

It may be easy to determine whether a person is stressed, but it is difficult to determine why. However, in this study, the in-house questionnaire aims to analyse stress in academic and abuse contexts without inducing abnormal stress. The collected data demonstrates a minor variation in the characteristics that differentiate academic stress from abusive stress. Different classifiers and modalities effectively distinguished between academic and abuse stress, as demonstrated by the study's findings. The kappa statistic was also calculated and compared; it was 0.64 for the best classifying model, demonstrating the classifier's ability to predict the classes from the provided dataset accurately.

This research can be used to identify stressed schoolgirls. Numerous studies [2], [13], [67] recommend that the university administration provide undergraduate and graduate students with stress awareness, stress management guidelines, self-guided programmes, and periodic counselling. Because schoolgirls are young and must be attentively supervised, it is crucial to adhere to such regulations. Consequently, schoolteachers, assigned authorities, and even parents can periodically assess the stress levels of females with the aid of this research. Any affordable heart rate measuring device and blood pressure monitor, which can be found in the majority of our residences today, can be used in schools and at home with medical advice. Using the developed machine learning model, one can determine whether a schoolgirl is stressed based on her behaviour, heart rate, and blood pressure, even if she is restrained from expressing herself.

Due to the delicate nature of the research, which involves themes such as psychological assessment and the stress experienced by young girls due to abuse or harassment, the schools that the authors approached to gather data were hesitant to give permission. Due to the school not being the designated study region, the researcher faced challenges in visiting each student's residence and acquiring parental consent to carry out the operation. Despite being the recipients of this research, several parents were reluctant to let their children's participation. Therefore, acquiring consent from participants and parents required a significant amount of effort. As a result, it contains information solely on fifty schoolgirls. The participants had diverse educational backgrounds, attended various schools, spanned different age groups, and hailed from different places. However, it is important to note that the research was limited to a few specific areas and did not encompass a broader range of people. Hence, the model's efficacy may have been enhanced by incorporating a greater number of samples encompassing diverse geographies and backgrounds.

5. Conclusions and Future Works

This research has collected data to examine the abuse that schoolgirls experience as a stressor that impacts their lives. This data would help identify schoolgirls who conceal abuse and stress-related issues. The research utilised simple and effective wearable devices to collect valuable data by differentiating between abuse and academic stress. With the in-house questionnaire, the research has also emulated the psychiatrist's technique of chatting with and monitoring the behaviour of the individuals. Multinomial LR was chosen as the best classifier due to its 89.9% F1 score, 89.9% precision, 90% recall, 88.5% ROC, and 65% kappa statistic. The decision tree outperforms others in terms of sensitivity, and each classifier contributes to the highest specificity value of each class. The acceptability of the study procedures and the comfortability of the wearable devices were assessed based on the participant's feedback, and they were 100%.

While the model attains a commendable level of accuracy, it can be further enhanced to achieve generalizability by gathering data from a wider range of individuals with diverse backgrounds and from other locations as a future endeavour. Conducting a more comprehensive evaluation of the participants' psychological state over several months can provide more dependable data, which can be utilised in future research. The utilisation of sophisticated machine learning algorithms and deep learning approaches can be employed to produce a more optimised model. Furthermore, the acquired data can be used to evaluate the intervention tactics that contribute to increased stress levels among schoolgirls.

Statements and Declarations

Author Contribution

DG, MJ, and PC contributed to the conception and design of the research protocol. DG performed the data collection. All authors contributed to the analysis and implementation of the work. The manuscript was written by DG, and other authors reviewed and commented on it. DG, MJ, and PC read and approved the final manuscript.

Funding

This research received no support from funding agencies.

Ethical Declaration

Informed consent was received from the schoolgirls and their parents. The research was conducted by the institutional research committee's ethical guidelines.

Acknowledgements

We want to thank Dr. A. Venkatesan, Psychiatrist, and Dr. M. Vinoth Kumar, Assistant Professor in Psychology, for sharing their knowledge in developing the questionnaire and comprehending the participants' gestures. We also thank the participants and their parents for their cooperation in completing the data collection.

Data Availability

The study's data is the property of Lincoln University College in Malaysia, and it will be accessible after a short period.

Conflicts of Interest

The authors report that there are no conflicts of Interest.

REFERENCES

[1] Congresses, "Stress and Mental Illness," *The Lancert*, vol. 2, no. 7039, pp. 205–208, 1958.

[2] C. Dunkel-Shetter and M. Lobel, "Stress Among Students," *New Dir. Student Serv.*, no. 49, pp. 17–34, 1990.

[3] N. Schneiderman, G. Ironson, and S. D. Siegel, "Stress and health: Psychological, behavioral, and biological determinants," *Annual Review of Clinical Psychology*, vol. 1. pp. 607–628, 2005. doi: 10.1146/annurev.clinpsy.1.102803.144141.

[4] S. D. McMahon, K. E. Grant, B. E. Compas, A. E. Thurm, and S. Ey, "Stress and psychopathology in children and adolescents: Is there evidence of specificity?," *J. Child Psychol. Psychiatry Allied Discip.*, vol. 44, no. 1, pp. 107–133, 2003, doi: 10.1111/1469-7610.00105.

[5] S. Inamdar, R. Chapekar, S. Gite, and B. Pradhan, "Machine Learning Driven Mental Stress Detection on Reddit Posts Using Natural Language Processing," *Human-Centric Intell. Syst.*, 2023, doi: 10.1007/s44230-023-00020-8.

[6] P. T. H. Nguyen, D. M. Le, V. Van Bui, S. Q. Le, and L. A. Thi Le, "The Reasons for Anxiety among Female University Employees," *Universal Journal of Public Health*, vol. 10, no. 5, pp. 519–526, 2022, doi: 10.13189/ujph.2022.100509.

[7] J. H. B. De Groot, M. A. M. Smeets, and G. R. Semin,

"Rapid stress system drives chemical transfer of fear from sender to receiver," *PLoS One*, vol. 10, no. 2, pp. 1–22, 2015, doi: 10.1371/journal.pone.0118211.

[8] Editorial, "The science of stress relief," *eBioMedicine*, vol. 78, pp. 1–2, 2022, doi: 10.1016/j.ebiom.2022.104018.

[9] D. Bhargava and H. Trivedi, "A Study of Causes of Stress and Stress Management among Youth," *Int. J. Manag. Soc. Sci.*, vol. 11, no. 03, pp. 108–117, 2018, doi: http://dx.doi.org/10.21013/jmss.v11.n3.p1.

[10] R. Aan, S. Ar, K. Aj, R. Chan, Z. Mi, and R. A. Al-naggar, "Stress and Coping Strategies among Management and Science University Students : A Qualitative Study," *Int. Med. J. Strateg.*, vol. 8, no. 2, pp. 11–16, 2009, doi: 10.31436/imjm.v8i2.751.

[11] K. G. Gomathi, S. Ahmed, and J. Sreedharan, "Causes of Stress and Coping Strategies Adopted by Undergraduate Health Professions Students in a University in the United Arab Emirates," *Sultan Qaboos Univ. Med. J.*, vol. 13, no. 3, pp. 1–9, 2013.

[12] S. George and B. B. Joseph, "Level of stress and its causes among 1 st -year dental students - A cross-sectional study," *Natl. J. Physiol. Pharm. Pharmacol.*, vol. 8, no. 11, pp. 1518–1521, 2018, doi: 10.5455/njppp.2018.8.0825017082 018.

[13] E. R. Schislyaeva and O. A. Saychenko, "The design of a mental health service platform for college students based on multi-modal information," *Microprocess. Microsyst.*, vol. 96, p. 104748, 2023, doi: https://doi.org/10.1016/j.micpro. 2022.104748.

[14] M. R. Febriansyah, Nicholas, R. Yunanda, and D. Suhartono, "Stress detection system for social media users," *Procedia Comput. Sci.*, vol. 216, pp. 672–681, 2023, doi: https://doi.org/10.1016/j.procs.2022.12.183.

[15] A. Ahmed *et al.*, "Machine learning models to detect anxiety and depression through social media: A scoping review," *Comput. Methods Programs Biomed. Updat.*, vol. 2, no. November 2021, p. 100066, 2022, doi: 10.1016/j.cmpbup.2022.100066.

[16] J. Katerine, G. Carrillo, and C. Manuel, "Academic stress detection on university students during COVID-19 outbreak by using an electronic nose and the galvanic skin response," *Biomed. Signal Process. Control*, vol. 68, no. March, 2021.

[17] J. Xu and X. Yang, "The influence of resilience on stress reaction of college students during COVID-19: the mediating role of coping style and positive adaptive response," *Curr. Psychol.*, vol. 19, no. 0123456789, 2023, doi: 10.1007/s12144-022-04214-4.

[18] B. Egilmez, E. Poyraz, W. Zhou, G. Memik, P. Dinda, and N. Alshurafa, "UStress: Understanding college student subjective stress using wrist-based passive sensing," *2017 IEEE Int. Conf. Pervasive Comput. Commun. Work. PerCom Work. 2017*, pp. 673–678, 2017, doi: 10.1109/PERCOMW.2017.7917644.

[19] J. Rodr ǵuez-arce *et al.*, "Towards an anxiety and stress recognition system for academic environments based on physiological features," *Comput. Methods Programs Biomed.*, vol. 190, 2020.

[20] N. Jaques, S. Taylor, A. Azaria, A. Ghandeharioun, A. Sano, and R. Picard, "Predicting students' happiness from physiology, phone, mobility, and behavioral data," in *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015*, 2015, pp. 222–228. doi: 10.1109/ACII.2015.7344575.

[21] X. Xu *et al.*, "Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–33, 2019, doi: 10.1145/3351274.

[22] B. Thanasekhar, N. Gomathy, K. A, and S. S, "Machine Learning Based Academic Stress Management System," in *2019 11th International Conference on Advanced Computing (ICoAC)*, 2019, pp. 147–151.

[23] X. Li, B. Hu, J. Shen, T. Xu, and M. Retcliffe, "Mild Depression Detection of College Students: an EEG-Based Solution with Free Viewing Tasks," *J. Med. Syst.*, vol. 39, no. 12, 2015, doi: 10.1007/s10916-015-0345-9.

[24] R. Ahuja and A. Banga, "Mental stress detection in university students using machine learning algorithms," *Procedia Comput. Sci.*, vol. 152, pp. 349–353, 2019, doi: 10.1016/j.procs.2019.05.007.

[25] D. Carneiro, P. Novais, D. Dur ães, J. M. Pego, and N. Sousa, "Predicting completion time in high-stakes exams," *Futur. Gener. Comput. Syst.*, vol. 92, pp. 549–559, 2019, doi: 10.1016/j.future.2018.01.061.

[26] L. V. Coutts, D. Plans, A. W. Brown, and J. Collomosse, "Deep learning with wearable based heart rate variability for prediction of mental and general health," *J. Biomed. Inform.*, vol. 112, no. February, 2020, doi: 10.1016/j.jbi.2020.103610.

[27] R. Castaldo *et al.*, "Detection of Mental Stress due to Oral Academic Examination via Ultra-short-term HRV Analysis," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 3805–3808.

[28] J. Xie, W. Wen, G. Liu, C. Chen, J. Zhang, and H. Liu, "Identifying strong stress and weak stress through blood volume pulse," *PIC 2016 - Proc. 2016 IEEE Int. Conf. Prog. Informatics Comput.*, pp. 179–182, 2017, doi: 10.1109/PIC.2016.7949490.

[29] Y. Wu, M. Daoudi, A. Amad, L. Sparrow, and F. D'Hondt, "Unsupervised learning method for exploring students' mental stress in medical simulation training," *ICMI 2020 Companion - Companion Publ. 2020 Int. Conf. Multimodal Interact.*, pp. 165–170, 2020, doi: 10.1145/3395035.34251 91.

[30] M. C. Pascoe, S. E. Hetrick, and A. G. Parker, "The impact of stress on students in secondary school and higher education," *International Journal of Adolescence and Youth*, vol. 25, no. 1. Routledge, pp. 104–112, 2020. doi: 10.1080/02673843.2019.1596823.

[31] L. J. Labrague, D. M. Mcenroe-petitte, D. Gloe, I. V Papathanasiou, and K. Tsaras, "A literature review on stress and coping strategies in nursing students," *J. Ment. Heal.*, 2016, doi: 10.1080/09638237.2016.1244721.

[32] V. S. Rathod and S. S. Sinha, "Prevalence Of Stress And Its Association With Resilience Among Medical Students In A Tertiary Care Center , Badnapur , Maharashtra : A Cross-

Sectional Study," *Eur. J. Mol. Clin. Med.*, vol. 10, no. 01, pp. 2753–2758, 2023.

[33] I. Ahmad, R. Gul, and M. Zeb, "A Qualitative Inquiry of University Student's Experiences of Exam Stress and Its Effect on Their Academic Performance," *Hum. Arenas*, 2022, doi: 10.1007/s42087-022-00285-8.

[34] C. Zhang *et al.*, "Associations Between Academic Stress and Depressive Symptoms Mediated by Anxiety Symptoms and Hopelessness Among Chinese College Students," *Psychol. Res. Behav. Manag.*, vol. 15, pp. 547–556, 2022, doi: 10.2147/PRBM.S353778.

[35] L. P. Jiménez-Mijangos, J. Rodríguez-Arce, R. Martínez-Méndez, and J. J. Reyes-Lagos, "Advances and challenges in the detection of academic stress and anxiety in the classroom: A literature review and recommendations," *Educ. Inf. Technol.*, pp. 3637–3666, 2022, doi: 10.1007/s10639-022-11324-w.

[36] Y. Deng *et al.*, "Family and Academic Stress and Their Impact on Students' Depression Level and Academic Performance," *Front. Psychiatry*, vol. 13, no. June, pp. 1–13, 2022, doi: 10.3389/fpsyt.2022.869337.

[37] J. Deng, L. Zhang, G. Cao, and H. Yin, "Effects of adolescent academic stress on sleep quality: Mediating effect of negative affect and moderating role of peer relationships," *Curr. Psychol.*, vol. 42, no. 6, pp. 4381–4390, 2023, doi: 10.1007/s12144-021-01803-7.

[38] A. Dhillon and A. Singh, "Machine Learning in Healthcare Data Analysis: A Survey," *J. Biol. Today's World*, vol. 8, no. 2, pp. 1–10, 2018, [Online]. Available: http://journals.lexispublisher.com/jbtw

[39] J. Mudoi, "a Study on Academic Stress Among Indian Teen-Age Girls," *J. Crit. Rev.*, vol. 06, no. 06, pp. 2674–2680, 2019.

[40] S. Ranasinghe, S. Ramesh, and K. H. Jacobsen, "Hygiene and mental health among middle school students in India and 11 other countries," *J. Infect. Public Health*, vol. 9, no. 4, pp. 429–435, 2016, doi: https://doi.org/10.1016/j.jiph.2015.11.007.

[41] CDC, "U.S. Teen Girls Experiencing Increased Sadness and Violence," 2023. https://www.cdc.gov/nchhstp/newsroom/2023/increased-sadness-and-violence-press-release.html

[42] Reuters, "Teen girls seeing 'dramatic' rise in poor mental health - U.S. CDC," 2023. https://www.reuters.com/world/us/more-high-school-girls-suffered-poor-mental-health-than-boys-us-data-2023-02-13/

[43] L. B. Chemutai and S. M. Mulambula, "The Determination of Academic Stress Levels by Gender and Year of Study. A Case of Kenyan University Students," *Am. Int. J. Soc. Sci.*, vol. 9, no. 4, pp. 16–22, 2020, doi: 10.30845/aijss.v9n4p3.

[44] C. Rubach, L. von Keyserlingk, S. D. Simpkins, and J. S. Eccles, "Does Instructional Quality Impact Male and Female University Students Differently? Focusing on Academic Stress, Academic Satisfaction, and Mental Health Impairment," *Front. Educ.*, vol. 7, no. February, pp. 1–16, 2022, doi: 10.3389/feduc.2022.820321.

[45] B. S. Graves, M. E. Hall, C. Dias-Karch, M. H. Haischer, and C. Apter, "Gender differences in perceived stress and coping among college students," *PLoS One*, vol. 16, no. 8 August, pp. 1–12, 2021, doi: 10.1371/journal.pone.0255634.

[46] W. Gao, S. Ping, and X. Liu, "Gender differences in depression, anxiety, and stress among college students: A longitudinal study from China," *J. Affect. Disord.*, vol. 263, pp. 292–300, 2020, doi: https://doi.org/10.1016/j.jad.2019.11.121.

[47] B. Chhetri, L. M. Goyal, M. Mittal, and G. Battineni, "Estimating the prevalence of stress among Indian students during the COVID-19 pandemic: A cross-sectional study from India," *J. Taibah Univ. Med. Sci.*, vol. 16, no. 2, pp. 260–267, 2021, doi: https://doi.org/10.1016/j.jtumed.2020.12.012.

[48] S. Tyagi and S. Karande, "Child sexual abuse in India: A wake-up call," *J. Postgr. Med.*, vol. 67, no. 3, pp. 125–129, 2021, doi: 10.4103/jpgm.JPGM_264_21.

[49] R. Ye, J. Li, Y. Du, H. Wang, and J. Gu, "Experience of childhood sexual violence and its associations with depressive symptoms among university students in Guangdong, China," *J. Affect. Disord.*, vol. 321, pp. 234–241, 2023, doi: https://doi.org/10.1016/j.jad.2022.10.039.

[50] S. Shoib *et al.*, "Relationship between childhood sexual abuse and suicidal behaviour in South Asian countries: A scoping review," *Psychiatry Res. Commun.*, vol. 2, no. 3, p. 100066, 2022, doi: 10.1016/j.psycom.2022.100066.

[51] E. N. B. Quarshie, "Self-Harm Among School-Going Adolescent Survivors of Sexual Violence Victimisation: A Cross-Sectional Study," *Front. Sociol.*, vol. 6, no. May, pp. 1–13, 2021, doi: 10.3389/fsoc.2021.605865.

[52] F. Leach and P. Gender, "The sexual harassment and abuse of adolescent schoolgirls in South The sexual harassment and abuse of adolescent schoolgirls in South India," *Educ. Citizsh. Soc. Justice*, vol. 2, no. 3, pp. 257–277, 2007, doi: 10.1177/1746197907081262.

[53] D. Incera and G. Manuel, "Homophobia is online : Sexual victimization and risks on the internet and mental health among bisexual , homosexual , pansexual , asexual , and queer adolescents," *Comput. Human Behav.*, vol. 119, pp. 1–10, 2021, doi:https://doi.org/10.1016/j.chb.2021.106728.

[54] J. S. Mainali P, Motiwala F, Trivedi C, Vadukapuram R, Mansuri Z, "Sexual Abuse and Its Impact on Suicidal Ideation and Attempts and Psychiatric Illness in Children and Adolescents With Posttraumatic Stress Disorder," *Prim Care Companion CNS Disord*, vol. 25, no. 1, 2023, doi: 10.4088/PCC.22m03239. PMID: 36705981.

[55] R. M. Furr and D. C. Funder, *Chapter 16- Behavioral Observation from Observing Behavior Psychology*. 2007.

[56] R. S. McGinnis *et al.*, "Rapid Anxiety and Depression Diagnosis in Young Children Enabled by Wearable Sensors and Machine Learning," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2018-July, no. March, pp. 3983–3986, 2018, doi: 10.1109/EMBC.2018.8513327.

[57] S. Swain and M. R. Patra, "Analysis of Depression, Anxiety, and Stress Chaos Among Children and Adolescents Using Machine Learning Algorithms," in *Innovations in Intelligent Computing and Communication*, 2022, pp. 261–272. doi: Ihttps://doi.org/10.1007/978-3-031-23233-6_20.

[58] R. Rois, M. Ray, A. Rahman, and S. K. Roy, "Prevalence

and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms," *J. Heal. Popul. Nutr.*, vol. 40, no. 1, pp. 1–12, 2021, doi: 10.1186/s41043-021-00276-5.

[59] H. Altaf, S. N. Ibrahim, and R. F. Olanrewaju, "Non Invasive Stress Detection Method Based on Discrete Wavelet Transform and Machine Learning Algorithms," in *2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 2021, pp. 106–111. doi: 10.1109/ISCAIE51753.2021.9431805.

[60] T. Nijhawan, G. Attigeri, and T. Ananthakrishna, "Stress detection using natural language processing and machine learning over social interactions," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00575-6.

[61] S Cohen, "Perceived Stress Scale," 1994. [Online]. Available: https://www.mindgarden.com/documents/PerceivedStressScale.pdf

[62] A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, *STOP, THAT and one hundred other sleep scales*. 2012. doi: 10.1007/978-1-4419-9893-4.

[63] F. Fazel, "Stress Level Test (Self- Assessment) Am I too stressed ? How much stress is too much ?" https://www.psycom.net/quizzes (accessed Feb. 22, 2023).

[64] P. Mavrenkova, N. Pankova, M. Lebedeva, and M. Karganov, "Heart Rate Variability Peculiarities in Adolescents with Psychopathology in the Educational Environment," *Universal Journal of Public Health*, vol. 9, no. 6, pp. 445–453, 2021, doi: 10.13189/ujph.2021.090612.

[65] Weka, "Waikato Environment for Knowledge Analysis." The University of Waikato, Hamilton, New Zealand.

[66] M. L. McHugh, "Lessons in biostatistics interrater reliability : the kappa statistic," *Biochem. Medica*, vol. 22, no. 3, pp. 276–282, 2012, [Online]. Available: https://hrcak.srce.hr/89395

[67] L. M. van der Lubbe, C. Gerritsen, M. C. A. Klein, R. F. Rodgers, and K. V. Hindriks, "Experiences of Users with an Online Self-Guided Mental Health Training Program Using Gamification," *J. Healthc. Informatics Res.*, 2023, doi: 10.1007/s41666-022-00124-z.